2

Preliminaries

This chapter introduces all fundamental concepts used in this research study.

2.1 · Multi-Objective and Many-Objective Optimization

Multi-objective optimization (also called multi-criteria optimization, multi-objective programming, multi-attribute optimization, vector optimization or Pareto optimization) is an area of decision making that is concerned with mathematical optimization problems involving more than one objective function to be minimized or maximized simultaneously. A multi-objective optimization (MOO) problem is defined by a number of objective functions $f_i: X \to \mathbb{R}$, $i = 1, ..., m$ to be maximized (or maximized) for some search space X. A solution $a \in X$ is said to dominate a solution $b \in X$, if and only if $\forall i$: $f_i(a) \leq f_i(b)$ and $\exists j : f_j(a) < f_j(b)$. Two solutions in $a \in X$ and $b \in X$ are non-dominated w.r.t. each other, if neither a dominates b nor b dominates a. The efficient set X_e is the set of solutions in X that is not dominated by any solution in X. The Pareto front PF is the image set of X_e, i.e. $PF = \{f(x) \mid x \in X_e\} \subseteq \mathbb{R}^m$.

Many-objective optimization (ManOO) applies to problems with more than three objective functions to optimize simultaneously [23]. There are new issues arising in many-objective optimization compared to multi-objective optimization. The two main issues in differentiating many-objective problems from multi-objective problems are the following: on the one hand, a large number of objective functions makes visualization of the Pareto front impractical; on the other hand, analysis of Pareto fronts is difficult due to the tendency that a majority of solutions will be non-dominated. Hence, the tradeoff analysis of conflicts between objective functions and the representation of the entire Pareto front can become difficult and in-transparent. Moreover, a high number

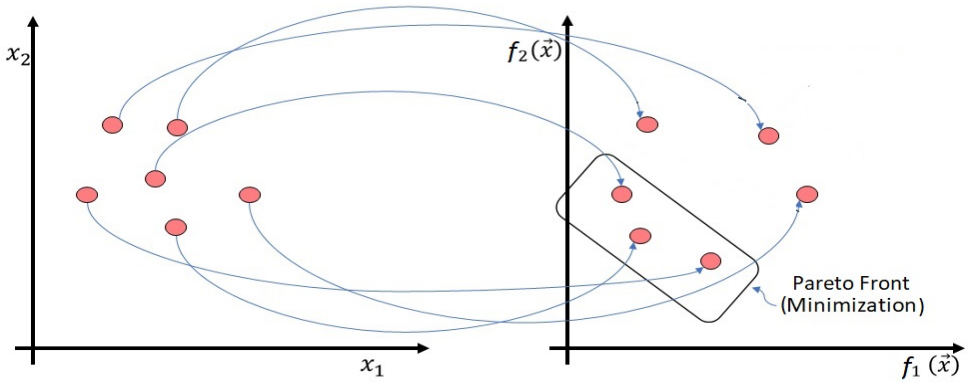

Figure 2.1 An illustration of multi-objective optimization for the decision variables and two objectives (to be minimized).

of objectives can cause to a significant increase in the computational time complexity required to compute Pareto fronts. In spite of these difficulties, many technique and approaches have been proposed to deal with many-objective optimization [33], [50], [40], [57].

2.2 · Networks

Network (or graphs) are one of the most fundamental data structure in computer science. A network can be represented by means of an adjacency matrix $A \in \mathbb{R}^{m \times m}$, where m denotes the number of nodes. Here the entry A_{ij} is zero if there is no connection between node i and node j, and non-zero otherwise. The non-zero number represents the weight of the connection. A network is used to represent the relationship between objects in a certain domain. An object in a network is a named node or vertex and the relationship between objects is called edge or link. The network can be used to describe a relationship between humans and their relationship in social life, countries in the world trading commodities, cities in a delivery problem in logistics, train stations or bus stops in some transportation system, connected computers in the Internet, airports in flight data set, interactions between proteins in biological system, and so forth. Analyzing such types of networks has become an immensely promising research area,and there is a lot of active research in network science, including community detection and network centrality.

2.2.1 · Community Detection

Community detection is a well-known method of social network analysis. A community is a group of nodes with many links between nodes of the group, but not so many links to nodes outside the group. The most popular community detection method for network clustering is the Louvain method, which is based on modularity maximization. Modularity is a concept that originates from social network analysis [42]. It is a quality measure or strength of partitioning of a graph into communities (partitions, groups or clusters). Maximizing modularity groups the nodes of a graph in such a way that intracluster graph distances (or edge weights) are minimized and inter-cluster graph distances (or edge weights) are maximized. Let A_{ij} denote the weight of the edge from node i to node j. Let m denote the number of the nodes, and $K_i{=}\sum A_{ij}$ denote the sum of weights of edges belonging to node i. Moreover, C_i is the community to which node i is assigned. Finally $\delta(.,.)$ is the Kronecker symbol, which is equal to 1, if and only if both arguments are equal to each other. Otherwise, it obtains the value of 0. Now the modularity is defined formally as:

$$Q_{signed} = \frac{1}{2m} \sum_{i,j \in \{1,\ldots,m\}} \left[A_{ij} - \left(\frac{K_i * K_j}{2m} \right) \right] \delta(C_i, C_j) \tag{2.1}$$

Modularity maximization is an NP-Hard problem. This can be shown by polynomial reduction of 3-PARTITION [19], and thus, in general, it is difficult to solve this problem by means of exact methods. There are, however, several fast heuristics available, such as the Louvain method [42], which is a greedy heuristic that finds high modularity partitions of a network in short time. The first phase of the Louvain method begins by placing each node in its own singleton 'community'. Then the looping over all nodes is done in the following way:

For each node i all neighbors, that is, all the nodes j such that A_{ij} is non-zero, are analyzed from the point of view of the gain computed after removing i from its community and placing it into the community of j. The node i is then put into the community for which the increase in modularity is largest. If none of the potential re-assignments of i into other communities is associated with a positive gain in modularity, i stays in its original community and the algorithm moves on to the next node. The loop

is repeated until no further improvements are obtained, i.e. when the modularity has reached a local optimum.
In the next phase of the algorithm, a new network is constructed with the communities of nodes obtained at the first phase of the Louvain method. The weights of the edges between the new nodes are given by the sum of the weights between all nodes between communities of the previous phase. When this phase is finished, a new phase is started, and so on. This creates a hierarchy of communities. The algorithm stops when a maximum of the modularity is obtained, or in practice when the last performed pass did not further increase modularity.

2.2.2 · Network Centrality

Network centrality is an important concept in network studies and analysis. As in everyday reality, a person, or organization in some way has the influence to generate some important decision for a community or even for a human being. Identifying an important person or organization can be recognized as the problem of identifying key players in a community. Many network centrality methods have been proposed to identify different key players in a social setting ([11], [12]) such as the following:

- Degree centrality, which focuses on the number of peers to which a node is connected [21].
- Betweenness centrality, which considers the number of shortest paths in the network that pass through a certain node [6].
- Closeness centrality, which measures distance from a certain node to all other nodes [43].
- Eigenvector centrality and PageRank, which consider the number of links from one node to other nodes, the importance of these nodes, and to how many these nodes themselves point to [10], [44], [49].

In this thesis, the emphasis will be more on eigenvector centrality. The reason behind this decision is that the method represents the most fundamental properties of centrality measures, and is a remarkably long studied method [52]. It is also very similar to the well known Google PageRank Methods.

Formally, the eigenvector centrality can be defined as follows: For a given graph G=(V, E) with $|V|$ being the number of nodes, let $A = (A_{v,t})$, $v \in \{1, ..., |V|\}$, $t \in \{1, ..., |m|\}$ be the adjacency matrix, i.e. $A_{v,t} = 1$ if node v is linked to node t, and $A_{v,t} = 0$ otherwise. The relative centrality score of node v is defined as

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t=1}^{|m|} A_{t,v} x_t$$

where $M(v)$ is a set of the neighbors of v and λ is an eigenvalue of A. In vector notation, the eigenvector centrality can be rewritten in a simple equation as $Ax = \lambda x$ and it becomes clear that x is an eigenvector of A and λ an eigenvalue. As there can be many eigenvectors of A, by convention, the eigenvector that corresponds to the biggest eigenvalue is considered. It consists of only positive components. There are two important factors that influence the eigenvector centrality of the node in the network. They are:

- The number of or total weight of links neighbors that point to the node.
- The centrality of neighbors that point to the node.

There is a possibility that nodes with more neighbors have a lower eigenvector centrality compared to nodes with fewer neighbors. This can happen because the neighbors of the less connected node have a higher centrality.

2.2.3 · Multiplex Networks

Multiplex networks are networks consisting of multiple edge sets for the same set of node. The network is made up of multiple layers, each of which represents a given operation mode. More clearly, it can be defined as graphs that consist of a number of, say n, nodes and m different edge sets for these nodes, called layers. The node set is denoted by V and the edge sets are denoted by E_l, $l \in \{1, ..., m\}$. A multiplex network is represented formally as $G = (G_1, G_2, \ldots, G_l, \ldots, G_m)$.

A visual illustration of the layers is shown in Figure 10. Here we assume that every network G_l is fully described by the adjacency matrix A_l with elements $A^l_{ij} = W^l_{ij} > 0$, where $A^l_{ij} = W^l_{ij} > 0$, if there is a link with a positive weight between nodes i and j in layer l, and $A^l_{ij} = 0$ otherwise.

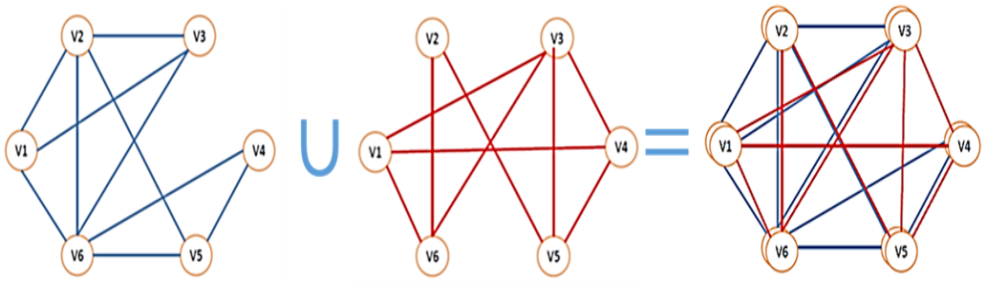

Figure 2.2 A visual illustration of multiplex network consisting of two layers of networks distinguished by blue and red colour. Each layer has different links but the nodes remain the same. The union of these layers, indicated by ∪, forms the multiplex network.

2.3 · Matrix Correlation Analysis

In statistics, the Pearson correlation coefficient, referred to as the Pearson's ρ, Pearson product-moment correlation coefficient (PPMCC) or bivariate correlation, [45] is a way to measure the linear correlation between two object. It has a value between -1 and 1, where 1 is perfect positive linear correlation, -1 is perfect negative linear correlation, and 0 is defined as no correlation among those two objects. This method was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s [20] [46] [53] and is widely used in the sciences, providing meaningful comparisons in system analysis.

We will now give the precise definition of how to compute the empirical correlation coefficient of two functions based on a finite number of evaluations. Let $(\Omega, \mathcal{S}, P)$ denote a probability space, where $\mathcal{S}$ is the event space and Ω denotes the set of elementary outputs – here chosen as the input space $X = \{0,1\}^N$. We will only consider singletons as events and write ω instead of $\{\omega\}$. In the following we consider the entire input space $X = \{0,1\}^N$, and a uniform distribution over this set. For each function F_i the random variables $\mathcal{F}_i : \Omega \to \mathbb{R}$ are defined as $\mathcal{F}_i : \omega \mapsto F_i(\omega)$. Next, consider a sample $\Omega' \subseteq \Omega$ and the realizations of random variables $\mathcal{F}_1(\omega), \ldots, \mathcal{F}_m(\omega)$ for $\omega \in \Omega$. Now, for the group of paired evaluations of $\mathcal{F}_i$ and $\mathcal{F}_j$ the empirical correlation coefficient can be computed as:

$$\rho_{ij}^{e} = \frac{\frac{1}{1-|\Omega_s|}\sum\limits_{\omega\in\Omega_s}(\mathcal{F}_i(\omega)-\overline{\mathcal{F}_i})(\mathcal{F}_j(\omega)-\overline{\mathcal{F}_j})}{\sqrt{\frac{1}{1-|\Omega_s|}\sum\limits_{\omega\in\Omega_s}(\mathcal{F}_i(\omega)-\overline{\mathcal{F}_i})^2}\sqrt{\frac{1}{1-|\Omega_s|}\sum\limits_{\omega\in\Omega_s}(\mathcal{F}_j(\omega)-\overline{\mathcal{F}_j})^2}}$$

As the matrix correlation applies to this research, it will utilize in the context of multi-objective and many-objective optimization, and decision making. The correlation can be interpreted as follows:

- Positively correlated objective functions can be interpreted as objective functions that support each other.
- Uncorrelated objective functions are considered to be independent of each other.
- Negatively correlated objective functions are in a strong conflict with each other.

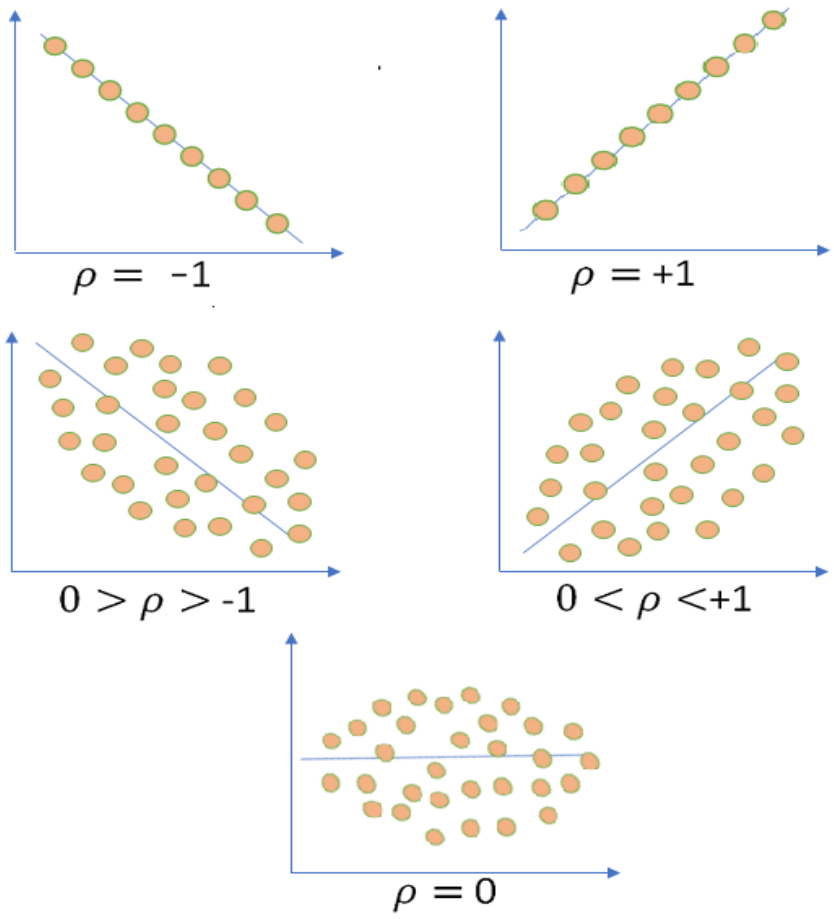

Figure 2.3 A visualization of the Pearson Correlation Coefficient and its range of values.